DOCUMENT RESUME

ED 271 480 TM 860 294

AUTHOR Royeen, Charlotte Brasic
TITLE A Comparison of Parametric versus Nonparametric Statistics.
PUB DATE Apr 86
NOTE 39p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Analysis of Variance; Data Analysis; Data Interpretation; *Nonparametric Statistics; Occupational Therapy; *Regression (Statistics); *Research Design; *Research Methodology; Research Needs; Research Problems; Statistical Studies
IDENTIFIERS *Parametric Analysis; Pearson Product Moment Correlation; T Test; *Univariate Analysis

ABSTRACT

In order to examine the possible effects of violation of assumptions using parametric procedures, this study is an exploratory investigation into the use of parametric versus nonparametric procedures using a multiple case study design. The case study investigation guidelines outlined by Yin served as the methodology. The following univariate statistical procedures, identified as the most commonly employed procedures in the occupational therapy literature (from 1980 to the present-time), served as the basis of the investigation: (1) independent t-test; (2) paired t-test; (3) one factor analysis of variance; (4) Pearson product moment correlation; and (5) regression. In each case investigated, meaningful patterns or characteristics of the data emerged which could aid interpretation regardless of class of data analysis procedure (parametric versus nonparametric) employed. Further, information on distributional characteristics of variables as revealed by exploratory data analysis compiled from research investigation over time served to define characteristics of variables studied in occupational therapy clinical research. Thus, it is recommended that occupational therapy clinical research include exploratory data analysis as a component for published research. Four pages of references and a discussion of the case study guidelines are appended. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED271480

A COMPARISON OF PARAMETRIC VERSUS NONPARMETRIC STATISTICS

A Paper Presentation to the Educational Statisticians Special Interest Group of the Annual Meeting of the American Educational Research Association

April 16, 1986

San Francisco, CA

Charlotte Brasic Royeen *

Office of Special Education Programs, U.S. Department of Education

MES 2313 3511, 400 Maryland Aveneue S.W., Washington, D.C.

* The author prepared this paper in her private capacity. No endorsement by the U.S. Department of Education is intended or should be inferred.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. Royeen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 860 294

BEST COPY AVAILABLE

Introduction

Data analysis procedures for clinical research are primarily centered upon the use of parametric statistical procedures which are based upon a normal distribution model. The assumption of a normal disribution is frequently justified using the central limit theorem (Edgington, 1969) which states that when randomly sampled, any population with finite variance will yield sample means that approach normality. However, the appropriateness of the central limit theorem for use with finite populations, as is often encountered in clinical research, is questionable.

To illustrate, Plane and Gordon (1982) state, "...when sampling from a finite population, increased sample size will not always bring the sample mean closer to normality" (p. 175). In fact, they found that sometimes the "...sampling distribution departs from nomality as the sample size approaches the finite population" (p. 175). Thus assuming the central limit theorem applies to all samples may not be prudent, and may lead to violation of one or more of the assumptions underlying use of parametric statistical tests.

Thus, the prevalent use of parametric procedures in clinical research may be questionable due to distributional characteristics of the data sets (Hill & Dixon, 1982). It is suggested by Lezak and Gray (1984) that"...the research data in clinical neuropsychology do not conform to the requirements for parametric statistical analysis" (p. 101). Lezak and Gray cite three primary reasons why parametric procedures are problematic in clincial neuropsychology and these may apply to clinical

BEST COPY AVAILABLE

research in general.

First, data sets in clinical neuropsychology are often fraught with outliers. Second, since the research studies are primarily clinically based, large samples are difficult to obtain due, in part, to mortality and morbidity. Third, there is an excessive amount of empty cells due to the wide range of variablility in behavioral dimensions and that given an array of tests, not all subjects can perform all tests and, consequently, there is considerable missing data.

As a solution to these problems, Lezak and Gray propose the increased use of nonparametric statistical procedures. Furthmore, McSweeney and Katz (1978) present an excellent summary of reasons for using nonparametric statistics. These reasons are (a) nonparametrics are assumption freer; (b) the data may not appropriately be put into metric form, i.e., unordered qualitative variables; (c) the data may be rank ordered; (d) the data may be from small samples; (e) there may be a non-normal distribution of the variables; (f) heterogeneity of variance between groups may exists; and, (g) outliers may be present.

Why are parametric tests so prevalent if, indeed, nonparametric methods offer more valid procedures when assumptions for parametric statistics are violated? One reason relates to power. Given a comparison between an equivalent parametric and nonparametric test run under conditions of normality, a parametric test has greater power (Siegel, 1956). Therefore, it is often assumed that a parmetric test is better.

BEST COPY AVAILABLE

But, how powerful is a parametric test when the assumptions upon which it is based are violated? Gibbons (1976) infers that when assumptions underlying parametric tests are violated, or if one cannot document that the assumptions underlying parametric statistics have been met, the concept of power is irrelevant since the tests may not be valid and the significance levels not accurate.

Another reason parametric statistics are more prevalent than nonparametric statistics has to do with the concept of robustness. Many parametric procedures, especially the F and t tests, are thought to be robust to violations of assumptions under certain conditions (Boneau, 1960, 1962; Box, 1953; Gardner, 1975; Minium, 1978). However, the claim of robustness to violations of assumptions appears to have been overgenerablized beyond highly specified situations (ususally delineated from simulation studies) to almost all real life situations and data sets. Thus, almost all violations of asssumptions, often ill defined, are tolerated in the name of robustness (Bradley, 1978; Singer, 1979). Furthermore, Bradley (1978) argues that there is no quantitative definition of robustness and that it is therefore a subjective claim and highly variable in form as it is commonly used.

Statement of the Problem

In order to study the possible effects of violation of assumptions using parametric procedures, the current study will be an exploratory investigation into the use of parametric versus

BEST COPY AVAILABLE

nonparametric procedures using a multiple case study design. That is, actual data sets within a given discipline - occupational therapy - will serve as units of analysis embedded in each "case" or published research.

The central issues to this investigation are (1) whether or not assumptions have inded been violated and (2) do violations of the assumptions make a difference to the substantive meaning of the staitistical analysis.

METHODS AND PROCEDURES

Overview

The case study process of investigation as outlined by Yin (1984; 1985) serves as the methodology for the current investigation. A case study methodology appropriately structures the investigation in order to "...understand events as they occur in their natural context" (Greene & David, 1984 p. 73).

Data Collection

Data Identification

All major articles published in the American Journal of Occupational Therapy, Occupational Therapy Journal of Research, Occupational Therapy in Mental Health, Physical and Occupational Therapy in Pediatrics, and Physcial and Occupational Therapy in

BEST COPY AVAILABLE

Geriatrics from 1980, or the onset of the journal, were indexed accroding to whether or not an article was statistically based. All of the statistically based articles were further reviewed for types of staitstical procedure employed. Inter-rater reliabillity was computed in a manner similar to that employed by Ottenbacher and Peterson (1985) and Goodwin and Goodwin (1985). In the current investigation the percentage of agreement score was 90%, which is comparable to that reported by Ottenbacher and Peterson (1985) and Goodwin and Goodwin (1985).

In order to appropriately limit the scope of the investigatiion, only univariate statistical procedures identified as the most commonly employed procedures in occupational therapy literature were studied. Thus, the following univariate procedures were identified as the most common in the occupational therapy literature and served as the basis of the investigation: (1) Independent t-test, (2) Paired t-test, (3) One factor ANOVA, (4) Pearson Product Moment Correlation, and (5) Regression.

Data Acquitision

Sample selection for case study investigations often utilized purposive sampling. That is, cases are not selected at random but are selected purposefully according to some theoretical rationale (Yin, 1984). Since Minium (1978) stated that nonparametric procedures are most appropriate to use when the sample size is small, data sets from published occupational therapy research were identified as possible cases for the current investigation when (a) one of the previously mentioned

BEST COPY AVAILABLE

parametric procedures was employed, and (b) when the sample size was small (less than fifty subjects). Accordingly, four to five articles per statistical procedure were identified for possible use. Of those articles so identified for each category of parametric procedure, the two with the smallest sample size were selected for possible use in the current study. Once such articles were identified, the authors were contacted by letter followed by telephone contact.

In all, five authors could not be located and contacted. Only one author refused to participate in the study, and three authors were willing to participate but could not locate their data sets. In cases of duplicative data sets for one category of procedure, the data set with the smallest sample size was used.

Data Analysis

Data Exploration

One or more of the following procedures was conducted with each data set in order to test the assumptions underlying use of the respective parametric analysis employed. All tests were performed on SAS (Statistical Analysis System) unless otherwise specified. The procedures used were:

(1) boxplots;

(2) the Shapiro Wilks test or W statistic;

(3) the F test was used to test for equality of variance in the two sample situations;

BEST COPY AVAILABLE

(4) the Barlett-Box F test of homogeneity of variance was used for three or more samples;

(5) the Squared Ranks test for distribution free testing of variance of k groups was computed by hand for appropriate cases (Conover, 1980);

(6) correlations of one variable with another, and;

(7) scattergrams.

Note that the above procedures refer to univariate tests for normality. In the bivariate case, the pearson Product Moment Correlation, procedures for testing bivariate normality were based upon Filliben (1975).

If exploratory data analysis could not be conducted due to small sample size, certain assumptions about the data set were made. First, it waas assumed that the central limit theorem did not necessarily apply. And, since evidence suggests that extremely small samples often do not follow a normal distribution (Lezak & Gray, 1984) it was assumed that disributional characteristics of the data set were questionable as regards normality.

Replication of Parametric Analysis and Selection and Calculation of Nonparametric Analysis

Each of the parametric analyses were replicated. Subsequently, the results of the exploratory data analysis were used in order to determine the appropriate nonparametric statistic. The nonparametric statistic was selected based upon the question the original author posed as well as guidelines

BEST COPY AVAILABLE

presented in Royeen and Seaver (1986).

Specific guidelines for procedures and interpretation regarding the comparison of parametric versus nonparametric statistics are included in the appendix.

Results and Discussion

Table 1 presents the results of the individual cases in terms of a multiple case study.

BEST COPY AVAILABLE

Table 1

Multiple Case Study Summary Table of Results of Parametric versus Nonparametric Statistics Using Data Sets from Occupational Therapy Literature

Parametric	Nonparametric	Conclusion
		CASE ONE
Paired t-test	Sign Test	Meaningful difference in p values. The nonparametric p value was much larger and better reflected the degree of similarity between the data sets. Skewness, two mild outliers and marginal normality were found in the data set of difference scores.
		CASE TWO
Independent t-test		
Common Dispersion	Mann Whitney Wilcoxon	No meaningful or substantive differences in p values for either comparison. One data set (strabismic sample) was mildly, positively skewed.
Uncommon Dispersion	Modified Mood's Median Test	
		CASE THREE
ANOVA Psychiatric	F test on Ranks	No meaningful or substantive differences in p values. Data of one group was a non-normal, uniform distribution and homogeneity of variance was marginal.
ANOVA Grade Point Average	F test on Ranks	No substantive differences in p values. A meaningful difference was not present using set criteria but results were close to a meaningful difference. Accommodator group of the data sets was mildly, negatively skewed with one mild and one extreme outlier.

Table 1 (continued)

Parametric	Nonparametric	Conclusion
ANOVA Physical Dysfunction	F test on Ranks	Substantive difference in p values. Reject null using nonparametric test, accept null using parametric test. Accommodator group of data sets was not normally distributed and heterogeneity of variance existed.
	CASE FOUR	
Pearson r	Kendall Tau	No substantive difference in p values. Meaningful difference with one set of paired variables (H). H set of variables was not bivariate normal and had two bivariate outliers.
Normal Children (n=15)		The nonparmetric p value is much less than the parametric p value.
Pearson r	Kendall Tau	Nine of the 14 correlations had substantive or meaningful (or both) differences in p values. In each instance the Kendall Tau was more conservative, yielding a larger p value. Univariate and bivariate abnormalities had been noted in all single and paired variables having p value differences. All of the nine cases of differences had bivariate outliers and lacked bivariate normality.
Learning Disabled (n=15)		
	CASE FIVE	
Normal Boys		
Parametric Regression	Nonparametric Regression	No meaningful or substantive difference in p values for slope or overall equation. Variable age had one mild outlier. Residual analysis for the normal boys data set yielded one outlier for the parametric analysis and mild skewness for the nonparametric analysis.

Table 1 (continued)

Parametric	Nonparametric	Conclusion
Normal Girls		
Parametric Regression	Nonparametric Regression	No meaningful or substantive difference in p values for slope or overall equation. Neither residual analysis yielded abnormalities.
Boys with Sensory Integrative Dysfunction		
Parametric Regression	Nonparametric Regression	No meaningful or substantive differences in p values for slope or overall equation. The variable age had one mild outlier. The results of the residual analysis yielded no abnormalities.

Table 1 reveals that in certain cases, and in certain instances within certain cases, there were clear differences in results between the parametric and the nonparametric statistics. However, in other cases there were no differences in results between the parametric and nonparametric statistics.

Research question one focused upon whether or not assumptions underlying the use of parametric statistics were met. Table 2 presents a summary analysis of assumptions underlying each case.

BEST COPY AVAILABLE

Table 2

Analysis of Assumptions Underlying Parametric Test in Each Case

Case	Procedure	Was Underlying Assumption Met?
1	Paired	No. Data of difference scores tested as marginally normal, had two mild outliers, and was mildly skewed.
2	Independent t-test	Yes, with a minor qualification. Data sets tested as normal but strabismic data set was mildly, positively skewed.
3	One factor ANOVA	No. True interval or ratio level of measurement was not employed. Samples did not appear to be from a normally distributed population. Heterogeneity of variance existed.
4	Pearson r	No. Bivariate normality lacking in nine (normal subgroup), and thirteen (learning disabled subgroup) of the paired variables.
5	Regression	Yes.

In only one case (regression analysis) were assumptions unequivocally and clearly met. In one case (independent t-test), assumptions were met with a minor qualification and in three cases assumptions clearly were not met. Thus, research question one can be answered that in three out of five cases assumptions underlying use of parametric procedures were not met during a multiple case investigation of small sample size data sets.

Thus, the belief that there is no need to test for

assumptions when using small data sets can be questioned due to the prevalence of violated assumptions with three of the five cases, purposefully selected due to small sample size.

Research question two, i.e., how similar or different are findings of parametric versus nonparametric tests, determined (a) whether violations of assumptions in these cases employing small data sets made a difference.

Table 3 is presented as a summary of interpretations pertaining to research question two.

BEST COPY AVAILABLE

Table 3

Parametric versus Nonparametric Findings for Individual Cases

Case	Procedure	Findings	Difference Defined
1	Paired t-test	Difference	Parametric p value is .5061. Nonparametric p value is .8146 to 1.0
2	Independent t-test	None	NA
3	ANOVA	Difference in one of three instances	Parametric p value is .0557. Nonparametric p value is .0297
4	Pearson r	One out of 14 is different for normal subgroup	Nonparametric p value is smaller than parametric (.09 vs. .24)
		9 out of 14 are different for learning disabled	Nonparametric p value is more conservative in every case
5	Regression	None	NA

Table 3 reveals that in cases one, three and four, the violated assumptions were indeed important. Furthermore, Table 3 reveals that the marginal normality and outliers present in the difference data set of case one apparently did affect the parametric test, and the outliers may have functioned like a 'heavy tail' which compromised the robustness of the paired t-test as had been indicated in the literature.

Table 4 presents a concise summary of these findings

BEST COPY AVAILABLE

regarding differences found related to whether or not parametric assumptions were met.

Table 4

Summary of Assumptions met Related to Differences Found By Case

Case	Procedure	Assumptions Met	Differences Found
1	Paired t-test	No	Yes
2	Independent t-test	Yes	No
3	One Factor ANOVA	No	Yes
4	Pearson r	No	Yes
5	Regression	Yes	No

Review of Table 4 reveals that in each of the two cases where parametric assumptions were met, no differences between parametric versus nonparametric procedures were found. And conversely, in each of the three cases where parametric assumptions were not met, differences between the parametric and nonparametric procedures were found. This finding indicates that if cases are considered as a whole, there is a one hundred percent agreement between whether or not parametric assumptions

BEST COPY AVAILABLE

were violated and whether or not differences were discovered regarding parametric versus nonparametric tests. This is support for the notion that assumptions need to be tested and selection of parametric or nonparametric data analysis procedures conducted accordingly.

Further analysis regarding research question three - how similar or different are findings - is based upon the concept of power. The use of parametric analysis in these five cases did not support the notion that when conditions underlying appropriate use of the parametric procedures are met, parametric tests are more powerful. Siegel (1956) and Fergusen (1976) had found parametric and nonparametric procedures to be within five percent range of comparable efficiency or power. Two cases within this multiple case study were found to be within an eleven percent comparable efficiency or power range: the independent t-test (case two) and regression with the subgroups normal girls, normal boys and boys with sensory integrative dysfunction (case three). For these cases, the nonparametric procedures appeared as powerful as the parametric procedure.

Thus, the claims that nonparametric procedures throw out information (Ager & Jacobson, 1980) and that they lack power (Boneau, 1960) are questioned since nonparametric procedures rendered equivalent findings to those from parametric procedures in these cases. In addition, David and Perez (1960) propose that the argument for using a nonparametric procedure is strengthened by knowing it will usually yield the same answer as a parametric test. The current multiple case lends support to this concept since, generally speaking, if assumptions weren't violated, the

BEST COPY AVAILABLE

nonparametric test performed comparable to the parametric test.

In certain instances when assumptions were violated, the nonparametric test was more powerful that the parametric test. This is consistant with Blair (1980) who stated that nonparametric tests can be more powerful when assumptions are violated. In the case of the physical dysfunction ANOVA (assumptions violated were normality and homogeneity of variance) the nonparametric p value was much smaller than the parametric value, suggesting the nonparametric test was more powerful to reject the null hypothesis in these circumstances. Use of the nonparametric test in this instance theoretically could have prevented a Type II error.

In another instance, use of the nonparametric test yielded a p value much closer to a level of significance than did the parametric procedure (Pearson r=.24, Kendall Tau=.09) when bivariate normality was lacking and there were two bivariate outliers. Again, the nonparametric test appeared more "powerful" in this case.

However, use of parametric tests when assumptions were violated also appeared to lead to Type I errors, that is innappropriate findings of significance. Thus, in these cases the nonparametric tests appeared to be more appropriate, not less powerful. For example, this happened in the case of the Pearson r with the subgroup of learning disabled (case four) for nine of the fourteen correlations.

Based upon the findings regarding Type I and Type II errors, it may be most appropriate not to define nonparametric procedures

BEST COPY AVAILABLE

as more or less powerful related to parametric procedures, but as possibly more validly employed with greater confidence in the Type I or Type II error rate probability.

Though not addressed by the research questions posited for the study, three issues are in need of discussion based upon the findings. These are (a) interactions of violations of assumptions, (b) distributional characteristics of the variables studied in occupational therapy clinical research, and (c) research design/methods in occupational therapy clinical research. Each of these will be discussed in turn.

Bradley (1978) noted that most simulation studies into robustness and effects of the violation of assumptions on parametric tests investigated the effects of single violations in isolation from each other. It is worth noting that in these authentic data sets all violations occurred in multiples of two or three. That is, it was never a single violation influencing the robustness of a particular test but rather the combination of two or three violated assumptions. These preliminary findings may have implications for continued research into the robustness of procedures when dual or multiple violations of assumptions are introduced into computer generated data sets for simulation studies. Furthermore, the concept of robustness as generalized from simulation research to practice may need re-evaluation in terms of the complexity of interactions between violated assumptions and the parametric tests as occurs in actual practice.

The distributional characteristics of the variables investigated by the studies comprising this multiple case

investigation are noteworthy regarding two dimensions; non-normality and outliers. In three of the five cases (paired t-test, one factor ANOVA and Pearson r) instances of univariate or bivariate non-normality were documented. These findings lend support to the notion proposed by Pocock (1982) that the central limit theorem may not apply in all sampling situations and that an automatic assumption of normality based upon the central limit theorem may not be prudent.

Furthermore, in four of the five cases (paired t-test, one factor ANOVA, Pearson r, and regression) there were outliers either in the original data set, the derived data set or calculated residuals. Lezak and Gray (1984) purport that data sets in clinical neuropsychology are often fraught with outliers, and it appears that - at least in these cases - the same is true for data sets in clinical occupational therapy research. Given the prevalence of outliers, the issue of "why" is paramount. Often, outliers are considered to be "errors" in a data set. Comrey (1985) states:

> These "bad" observations can result from many different causes, e.g., errors in test scoring, recording, and/or keypunching; errors in reading data into the computer, such as incorrect formatting, extra or missing cards or lines in the data deck or data set; computer errors; errors in observation or recording; deliberate falsification of responses by subjects for any one of a variety of reasons, and so on. (p. 273)

At this point, it is difficult to assess whether (a) the prevalence of outliers is an artifact of small data sets such as employed in this multiple case investigation, or (b) outliers represent valid manifestations of variables of interest in

BEST COPY AVAILABLE

occupational therapy clinical research. The outlier problem is not unique to occupational therapy. For example, Tupper and Rosenblood (1984) have discussed how characteristics of attribute variables can confound neurological research. Recently, the problem of outliers and non-normality was raised by Mooijaart (1985) regarding factor analytic procedures. He stated that, as regards non-normality and factor analysis:

> Unfortunately there are few psychological theories about how variables are distributed. It seems reasonable to apply models in which weak distributional assumptions are made. (p. 324)

Mooijaart's solution to the problem of non-normality and factor analysis seems reasonable to apply to occupational therapy clinical research. That is, until more research is done on distributional characteristics of variables of concern in occupational therapy clinical research, and unless variables under investigation are documented to be distributed normally, it appears prudent to apply data analysis procedures in which weak distributional assumptions are made, i.e., nonparametric procedures.

The third issue, occupational therapy research design and methods, is brought to light considering the state of the art of the application of design/methods in occupational therapy research over the period 1980 through 1984. For example, indexing of the occupational therapy literature from that time revealed that 66 percent of the data analysis procedures employed could be considered "basic" whereas only 31 percent were "advanced". The fact that the bulk of the data analysis procedures dealt with basic statistics suggests that occupational

BEST COPY AVAILABLE

therapy clinical research is in an emergence period and not yet fully developed in terms of design/methodological/statistical expertise. The field does not yet have a strong and vigorous methodology tradition.

The state of the art of the practice of research/design methods as illustrated by the purposive sample of procedures within this multiple case study further confirms the emergence of clinical research in occupational therapy, but highlights the lack of a rigorous tradition of methodology. To illustrate, two of the five cases within the multiple case study employed data analysis procedures which lack a reflection of advanced practice in research design/methods.

Thus, a confounding variable affecting the study of parametric versus nonparametric statistics using authentic data sets is design of the investigation from which data sets are acquired. For, it is difficult to evaluate similarities or differences between parametric verus nonparametric procedures when the research design is problematic. For example, the correlational investigation into test-retest reliability (case five) may have been more appropriately investigated using a research design and analysis procedures based upon analysis of variance (Cronback, 1947; Jackson & Messick, 1967). Similarly, the three one-factor ANOVA's executed (case three) may have been more appropriately executed as a factorial ANOVA with subsequent multiple comparisons (Box, Hunter & Hunter, 1978).

Consequently, it must be noted that the process of improving the quality and quantity of clinical research in occupational therapy will not be accomplished by a singular approach along one

BEST COPY AVAILABLE

dimension, i.e., parametric versus nonparametric procedures. As an illustration, Baum, Boyle and Edward (1984) state "The most difficult phase of clinical research is the "design . . .phase" (p. 267). And, it is the domain of research design/methods (of which parametric versus nonparametric procedures is just a part) which needs to be addressed. In this vein, Box (1984) discusses the general importance of practice regarding development of statistical procedures. In occupational therapy clinical research, the relationship between practice, research design and statistical procedures is crucial and underlies the future development of a tradition of research methods and practice unique to occupational therapy clinical research. The adequate development of such may be predicated upon fostering and generating research methods within the profession of occupational therapy, which in turn may be predicated upon the level of training (bachelor's or master's level) for entry into the profession (Royeen, in press-a). For, it is impossible to facilitate the growth of a tradition of methodology in occupational therapy without support for graduate level education in methodology related to occupational therapy (master's and doctoral level).

Conclusions and Future Research

The current investigation highlights the importance of exploratory data analysis when analyzing clinical data. In each case investigated herein, meaningful patterns or characteristics of the data emerged which could aid interpretation regardless of

BEST COPY AVAILABLE

class of data analysis procedure (parametric versus nonparametric) employed. Furthermore, information on distributional characteristics of variables as revealed by exploratory data analysis compiled from research investigations over time could serve to define characteristics of variables studied in occupational therapy clinical research. Thus, it is recommended that occupational therapy clinical research include exploratory data analysis as a component for published research.

Research into parametric versus nonparametric procedures is, as stated by McSweeney and Katz (1978), "equivocal". Since the current multiple case investigation can be applied to nonparametric theory directing application and applied only as individual cases resemble the cases presented herein, continued research which could be generalizable is warranted. Such future research could be structured as follows.

First, based upon design and methodology developed for the current investigation, conduct of multiple case investigations into single categories of procedures using data sets randomly selected from the occupational therapy literature from 1980 through 1986 can build upon the current investigation. The most commonly employed procedures in occupational therapy should be investigated, and these would be (in order of frequency in the literature): (a) Pearson r, (b) independent t-test, (c) factorial ANOVA and post hoc analyses, (d) one factor ANOVA, (e) pair t-test, (f) regression, and (g) ANCOVA. Second, future research could also address design/methodological issues and employ alternate designs (such as factorial ANOVA versus multiple one

BEST COPY AVAILABLE

factor ANOVA's and ANOVA based reliability analyses versus reliability analyses based upon correlations).

ACKNOWLEDGEMENTS

Preparation of the article was supported in part by a Doctoral Fellowship from the American Occupational Therapy Foundation. A copy of the dissertation upon which this article was based is available from Dissertation Abstracts International.

Special acknowledgement is given to those clinical researchers who graciously shared their data sets in the true spirit of science.

BEST COPY AVAILABLE

REFERENCES

Ager, C. L.. & Jacobson, B. R. (1980). Pure versus practical application of research methods: The robustness of the t and F. Letter to the Editor, American Journal of Occupational Therapy, 34(6), 406.

Baum, C.M., Boyle, M.A., & Edwards, D.F. (1984). Initiating occupational therapy clinical research. Amercian Journal of Occupationl Therapy, 38(4), 267-269.

Blair, R. C. (1981). A reaction to "Consequences of failure to meet assumptions underlying fixed effects analysis of variance and covariance". Review of Educational Research,51(4), 499-507.

Boneau, C. A. (1960). The effects of violations of the assumptions underlying the t test. Psychological Bulletin, 57(1), 49-64.

Boneau, C. A. (1962). A comparison of the power of the U and t tests. Psychological Review, 69, 246-256.

Box, G. E. P. (1984). The importance of practice in the development of statistics. Technometrics, 26(1), 1-8.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). Statistics for experimenters. New York: John Wiley and Sons.

Box, G. E. P. (1953). Non-normality and tests of variance. Biometrika, 40, 318-335.

Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, J. W. (1983). Graphical methods for data analysis. Boston: Duxbury Press.

Comrey, A. L. 91985). a method for removing outliers to improve factor analytic results. Multivariate Behavioral Research, 20(3), 273-280.

Conover, W. J. (1980). Practical Nonparametric Statistics. New York: Wiley and Sons.

Cronbach, L. (1947). Test reliability: It's meaning and determination. Psychometrika, 12, 1-16.

Cunninghan, S., & Trickey, B. (1983). The correlation of learning styles and student performance in academic and clinical coursework. Occupational Therapy Journal of Research, 3(1), 54-55.

David, H. A., & Perez, C. A. (1960). On comparing different tests of the same hypotheses. Biometrika, 47(3+4),

BEST COPY AVAILABLE

297-306.

Edgington, E. S. (1969). Statistical inference: The distribution free approach. New York: McGraw Hill.

Eisenhart, C., & Swed, F. S. (1960). On certain criteria for testing the homogeneity of variance. Annals of Mathematical Statistics, 11, 111-112.

Ferguson, F. A. (1976). Statistical analysis in psychology and education. New York: McGraw Hill.

Filliben, J. J. (1975). The probability plot correlation coefficient test for normality. Technometrics, 17, 1.

Gans, D. J. (1984). The search for significance: different tests on the same data. Journal of Statistics: Computer Simulation, 19,1-21.

Gardner, P. L. (1975). Scales and statistics. Reveiw of Educational Research. 45(1), 43-57.

Gibbons, J. D. (1985). Statistical inference. New York: Marcel Decker, Inc.

Gibbons, J. D. (1976). Nonparametric methods for quantitative analysis. Columbus, Ohio: American Sciences Press, Inc.

Goodwin, L. D., & Goodwin, W. L. (1985). Statistical techniques in American Educational Resarch Journal Articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Research, February, 5-11.

Greene, D., & David, D. L. (1984). A research design for generalizing from multiple case studies. Evaluation and Program Planning, Vol. 7, 73-85.

Hill, M. A., & Dixon, W. J. (1982). Robustness in real life: A study of clinical laboratory data. Biometrics, 38, 377-386.

Jackson, D. N., & Messick, S. (1967). Problems in human assessment. New York, McGraw Hill.

Leedy, P. D. (1974). Practical research planning and design. New York: MacMillan and Co.

Lezak, M. D., Gray, D. K. (1984). Sampling problems and nonparametric solutions in clinical neuropsychological research. Journal of Clinical Neuropsychology, 6(1), 101-109.

BEST COPY AVAILABLE

Marascuilo, L. A., & McSweeney, M. (1977). Nonparametric and distribution free methods for the social sciences. Montery, CA: Brooks-Cole Publishing Co.

McSweeny, M., & Katz, B. M. (1978). Nonparametric statistics: Use and nonuse. Perceptual and Motor Skills, 4(3), 1023-1032.

Minium, E. W. (1978). Statistical reasoning in psychology and education. New York: John Wiley and Sons.

Mooijaart, A. (1985). Factor analysis for non-normal variables. Psychometrika, 50(3), 323-342.

Ottenbacher, K. & Peterson, P. (1985). Quantitative trends in occupational therapy research: Implications for practice and education. American Journal of Occupational Therapy, 39(4),240-246.

Ottenbacher, K., Dauck, B. S., Givelinger, M., Grahm, V., & Hasset, C. (1985). Reliability of the behavioral assessment scale of oral functions in feeding. American Journal of Occupational Therapy, 39 (7), 436-441.

Plane, D. R., & Gordon, K. R. (1982). A simple proof of the nonapplicability of the central limit theorem to finite populations. American Statistician, 36(3), 175-176.

Pocock, S. J. (1982). When not to rely on the central limit theorem - an example from absenteeism data.

Royeen, C. B. (in press-a). Entry level education in occupational therapy. American Journal of Occupational Therapy

Royeen, C. B. (in press-b). The boxplot: A screening test for research data. American Journal of Occupational Therapy

Royeen, C. B., & Seaver, W. F. (1986). Promise in nonparametrics. American Journal of Occupational Therapy, 40(3), 191-193.

Siegel, S. (1956). Nonparametric statistics for the behavior sciences. New York: McGraw Hill.

Singer, B. (1979). Distribution-free methods for nonparametric problems: a classified and selected bibliography. British Journal of Mathematical and Statistical Psychology, 32, 1-60.

Statistical Analysis System User's Guide. (1982). Cary: North Carolina.

BEST COPY AVAILABLE

Tukey, J. W. (1977). Exploratory data analysis. Reading, MA: Addison Wesley Co.

Tupper, D. E., & Rosenblood, L. K. (1984). Methodological commentary: methodological considerations in the use of attribute variables in medical research. Journal of Clinical Neurology, 6(4), 441-453.

Velleman, P., & Hoaglin, D. C. (1981). Application, basics and computing of exploratory data analysis. Boston: Duxbury Press.

Walberg, H. J., Strykowski, B. F., Rovai, E., & Hung, S. S. (1984). Exceptional performance, Review of Educational Research, 54(1), 87-112.

Yerxa, E. J. (1984). Evaluation versus research: Outcomes or knowledge? American Journal of Occupational Therapy, 38(6), 407-408.

Yin, R. K. (1985). Case Study Methodology. Presentation to the Institute for Special Education Studies, Sponsored by the Division of Educational Services, U. S. Department of Education, November 6, Washington, D. C.

Yin, R. K. (1984). Case study research: Design and methods. Beverly Hills: Sage Publications

BEST COPY AVAILABLE

APPENDIX

Case Study Guidelines

Yin (1984) identified the delineation of the criteria for interpretation as a necessary step in case study methodology. Therefore, the following guidelines have been developed in order to structure the interpretation process of comparing parametric versus nonparametric tests. This process of interpretation is a complicated one based upon iterative analysis of the data sets (Chambers, Cleveland, Kleiner, and Tukey, 1983) and factors related to the statistical tests and assumptions underlying their use.

A limitation to the interpretation guidelines needs to be presented. In describing exploratory data analysis, Velleman and Hoaglin (1981) state:

> Exploratory methods . . . call for frequent application of the analysts' judgement, and the judgement cannot readily be cast in simple rules . . ." (p. XVI).

Indeed, this statement is applicable to the entire process of developing guidelines for the comparison of parametric versus nonparametric tests because situation specific difficulties can arise which may alter guidelines and demand judgmental decisions. However, considering this possible limitation to the guidelines, they are presented in six main categories: preliminary procedures, analysis of characteristics of the data set, analysis of characteristics of the statistical tests, comparison of significance levels, supporting data, and synthesis. The following are guidelines for the preliminary procedures.

BEST COPY AVAILABLE

Guidelines for Preliminary Procedures.

I. Research Design and Statistical Test

A. What is the research question expressed by the original author?

1. Specify the hypothesis.

2. State alternative hypothesis or outcomes.

B. What is the research design?

1. Identify number of groups in the study.

2. Identify characteristics of the groups regarding validity and reliability of measurement.

C. Confirm the parametric statistical test used by the author.

D. Does the test fit with the research design?

1. If yes, continue to E.

2. If no, evaluate suitability of data set for use in the study.

E. What are the conclusions of the author?

1. What is the statistical conclusion?

2. What is the substantive conclusion?

F. At what the level of measurement are the variables?

1. Use criteria of Leedy (1974).

2. Use criteria of Siegel (1956).

3. Use criteria of Minium (1978).

H. Is the level of measurement consistant with the statistical test employed?

1. If yes, continue to II.

BEST COPY AVAILABLE

2. If no, evaluate suitability of the data set for use in the current study.

II. Replication of the Parametric Analysis

A. Execute the parametric analysis.

B. Has the parametric analysis of the original author been replicated?

1. If yes, continue to C.

2. If no, investigate why. Contact original author to clarify situation as needed.

C. Review the assumptions underlying the use of the parametric test.

The following are the case analysis guidelines.

Case Analysis Guidelines.

I. Analysis of the Characteristics of the Data Set

A. What is the size of the data set.

1. If the data set or subset thereof has a sample size of five or more, that data set or subset thereof will be subjected to exploratory data analysis.

2. If the data set or subset thereof has a sample size of less than five, it will not be subjected to exploratory data analysis since the sample size is too small to render judgments about EDA.

B. What is the shape of the data set?

1. This will be revealed by a boxplot and a stem and leaf plot.

BEST COPY AVAILABLE

2. Is the data set symmetric?

3. Is it skewed? Skewness is determined by an absolute value greater than .5 (Walburg, Stykowski & Hung 1984). Mildly skewed would be a score of .5 to .99. Moderately skewed would be a score of 1.0 to 1.99. Profoundly skewed would be a score of 2.0 or greater.

4. Are there long tails present?

C. Is the distribution normal?

1. Univariate tests of normality such as the Shapiro-Wilkes will address this question. The normal probability plot will also provide information.

2. Bivariate normality will be determined by the process of computing the generalized distance squared and chi square values of each two variables upon which correlations are based. Consequently, bivariate normality will be determined considering the following:

2.1 Percentage of values falling within the ellipsoid at the .5 level. At least 50% of the scores must be at this level.

2.2 Percentage of values falling within the ellipsoid at the .9 level. At least 90% of the scores must be at this level.

2.3 Linearity of the scattergram of the chi square value by the generalized distance squared.

BEST COPY AVAILABLE

2.4 Degree of correlation between the generalized distance squared to the chi square value. An acceptable correlation is .90 or greater.

D. What is the range of the data?

1. The range will be revealed by stem and leaf displays, standard deviations, and min/max determinations.

2. Gaps in the data will be determined by the stem and leaf plot.

3. Repeated values will be revealed by the stem and leaf.

4. Single outliers will be revealed by the box plot.

5. Two or more data points acting together as an outlier will be revealed by the scattergram.

E. Is the relationship between variables linear?

1. Results revealed by scattergram.

2. If yes, continue to F.

3. If no, identify the nature of the relationship.

F. Using appropriate evidence drawn from the exploratory data analysis, conclusions about the distributional characteristics of the data sets will be summarized and justified.

II. Evaluation of Assumptions Underlying Use of the Parametric Test

A. Determination will be made of which assumptions were tested during the exploratory data analysis. Any

BEST COPY AVAILABLE

.118 will be used as this was the median difference identified by Gans (1984) when he investigated five different tests run on seven different samples varying by six distributions.

1. If the results between the parametric versus the nonparametric test are ambiguous, run a robust procedure.

    1.1 Select a robust procedure.

    1.2 Execute it.

    1.3 Analyze it.

    1.4 Go to interpretation.

2. If there was a clear difference between the results of the parametric and nonparametric analyses, but unknown distributional characteristics of the data set (N is less than five) run a robust procedure if feasible. If not, assume the ambiguity surrounding assumptions of the parametric procedure weighs in favor of the nonparametric procedure, and go to interpretation.

3. If there is a clear difference in the parametric versus the nonparametric test under known distributional assumptions and documented violations of assumptions, go to interpretation.

IV. Guidelines for Interpretation

A. Compare the probability levels considering ARE under normal conditions identified in Table 6 and the EDA. Also, refer to literature on power conditions of nonparametric tests under conditions of non-normality,

BEST COPY AVAILABLE

as available.

B. Compare the probability levels considering which, if any, assumptions have been violated and relate it to the literature.

C. Consider the substantive results between the sets of procedures in light of subjects and measurement characteristics.

D. Compare probability levels for those sets with robust procedures.

E. Consider any supporting or clarifying data from the literature.

VI. Synthesis

Based upon all interpretation, an overall synthesis will be made relative to which procedure, if either, was more appropriate in which case and why.

BEST COPY AVAILABLE

ACKNOWLEDGEMENTS

Appreciation is extended to the American Occupational Therapy Foundation for the financial support of a Doctoral Fellowship which, in part, funded this investigation. Complete copies of the dissertation upon which this is based can be obtained from University Microfilms, Inc., Ann Arbor, MI.

The clinical researchers who shared their data for the investigation are heartily thanked. They shared their data in the true spirit of science and represent commitment to the scientific process.